

語りの傾聴において不同意される発話とその検出

伊藤滉一郎¹ 村田匡輝² 大野誠寛³ 松原茂樹^{1,4}¹ 名古屋大学大学院 情報学研究科 ² 豊田工業高等専門学校³ 東京電機大学 未来科学部 ⁴ 名古屋大学 情報連携推進本部

ito.koichiro.v1@s.mail.nagoya-u.ac.jp murata@toyota-ct.ac.jp

ohno@mail.dendai.ac.jp matsubara.shigeki.z8@f.mail.nagoya-u.ac.jp

概要

コミュニケーションロボットやスマートスピーカーなどの会話エージェントが、人に代わって聴き手を担うことが期待されている。これらが聴き手として認められるには、傾聴を示す目的で語りに応答する発話（傾聴応答）の表出が有効である。傾聴では、語り手の発話に同意や肯定をすることが基本的な応答方略となる。しかし、語り手の自虐的な発話などには、語りに同意しないことを示す応答（不同意応答）が好ましい。本論文では、不同意応答の生成に向けて、その表出が適する発話の検出について述べる。不同意応答が適する発話が含まれる高齢者の語りデータを用いて、検出実験を実施し、その実現可能性を考察した。

1 はじめに

語ることは人間の基本的な欲求である。語る行為は、聴き手がいて初めて成立する。日本では、独居高齢者の増加など社会の個人化が進み [1], 聴き手不在の生活シーンが増加している。人が語れる機会を増やすことは現代社会の重要な課題である。これに対し、コミュニケーションロボットなどの会話エージェントやスマートスピーカーなどの情報機器が語りを聴く役割を担うことが考えられる。これらが聴き手として認められるには、語りを傾聴していることを語り手に伝達する機能を備える必要がある。このための明示的な手段は語りに応答することであり、ジェスチャや発話の表出が有力である。以降では、傾聴を示す目的で語りに応答する発話を**傾聴応答**と呼ぶ。

語りの傾聴では、語り手の発話に同意または肯定することが聴き手の基本的な応答方略となる。傾聴応答の代表は相槌であり、その生成法が提案されている [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]。しかし、語り手

表1 語りと傾聴応答の例

語り	傾聴応答
イタリア旅行をしたことが1番楽しかったです	はい（相槌） そうですかー（感心） 素敵ですねー（評価）
もう2度と行けないかなと思ってましたけど	イタリア旅行（繰り返し） いえいえ（不同意）

の自虐的な発話などには、同意や肯定を目的とした応答の表出が適さないことがある。その場合には、語りに同意しないことを示す応答の表出が求められる。以降では、こうした応答を**不同意応答**と呼ぶ。

本論文では、不同意応答の自動生成の実現に向けて、その表出が適する発話の検出について述べる。不同意応答が適する発話を含む高齢者の語りデータを用いた実験を行い、BERT [13] を用いた検出手法の性能を評価するとともに、その課題を考察した。

2 語りの傾聴と不同意

2.1 傾聴応答

傾聴応答は、語りを傾聴していることを語り手に伝える応答である。傾聴応答の種類には、相槌、感心、評価などがあり [14], 語り手の語る意欲を促進する効果がある。表1に、語りと傾聴応答の例を示す。()内は傾聴応答の種類を示す。

2.2 関連研究

傾聴応答の代表は相槌であり、その生成方法に関する研究が存在している [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]。これらの研究では、語りから抽出されるピッチやパワーなどの音響情報や、単語や品詞などの言語情報から、相槌の生成に適したタイミングを検出する手法を提案している。これまでに、ルールベースによる手法 [3, 7], n-gram モデルによる手法 [4], 有限状態トランジェューサによる手法 [5], 決定木に

よる手法 [2, 6], CRF による手法 [8], SVM による手法 [9, 10] などが提案されている。近年では, LSTM や BERT などのニューラルネットワークを用いた手法 [11, 12] も提案されている。

相槌以外の傾聴応答の生成に関する研究も存在する [15, 16, 17, 18, 19, 20, 21, 22, 23]。相槌だけでなく, 繰り返しや問い返しなども生成可能なシステムの開発に取り組まれている [15, 16, 17, 18]。これらのシステムは, 事前に定義したルールやアルゴリズムに従い, 生成する応答の種類を選択した後, 応答を生成する。他にも, 系列変換モデルによる, 傾聴応答の応答文の生成に関する研究も行われている [19]。また, 相槌以外の特定の種類の傾聴応答に着目した研究も存在する。繰り返し応答の生成に関しては, 繰り返しすべき語句の検出 [20, 21] や, 繰り返しの応答文の生成 [22, 23] が研究されている。従来の傾聴応答の生成に関する研究では, 語りの発話に同意や肯定を示す応答が, 主な生成の対象とされてきた。

2.3 傾聴における不同意応答

これまでに, Ito らは, 高齢者の語りに対する傾聴応答を収集している [24]。合計で約 15 万個の傾聴応答を収集しており, それらには人手で応答の種類が付与されている。応答の種類は, 相槌, 感心, 繰り返し, 評価, 同意など, 全 16 種類存在し, 全体の 67.96% が相槌であり, 相槌以外の種類は合計して 32.04% を占めていた。図 1 に, 相槌以外の応答の種類の分布を示す。相槌以外の応答の種類は, 感心, 繰り返し, 評価, 同意の順に多く出現していた。以下に, それらの応答と対応する語りの例を示す。

• 感心

【語り】 地方に行った時そこにある美術館にはなるべく行くようになっています

【応答】 そうなんです

• 繰り返し

【語り】 わたくしのこんにちまでの仕事はライターです

【応答】 ライター

• 評価

【語り】 書道も好きで総理大臣賞も頂いたりして

【応答】 凄いですね

• 同意

【語り】 五千歩歩くということはなかなか難しいことで

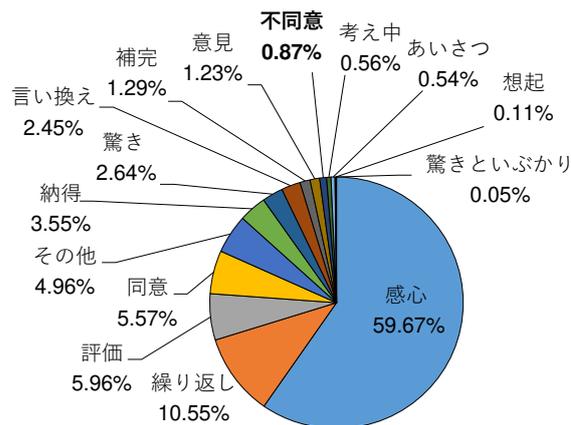


図 1 傾聴応答の種類の分布

【応答】 そうですね

語りを傾聴する聴き手の応答方略としては, 上述した例のように, 語りに同意や肯定を示す応答の表出が基本である。しかし, 場合によっては, 不同意応答の表出を必要とすることもある。例えば, 次に示すような自虐的・謙遜的な発話には, 同意や肯定を示す応答ではなく, 「いえいえ」などの不同意応答の方が適切である。

• <自虐・謙遜> 私は昔から気が利かないものでして

また, 次に示すような謝罪の発話も, 不同意応答の方が適切な語りの発話として挙げられる。

• <謝罪> 昔のことなので思い出せなくてごめんなさい

語りには, 上述したような不同意応答が適する発話の出現は, 多くはないものと考えられる。図 1 の分布においても, 不同意応答の出現は少数である。しかし, 不同意応答が適切な場面で, 誤って肯定や同意の応答をしてしまうと, 語り手に多大な不信感を与えてしまう。そこで, 本論文では, 不同意応答の自動生成の実現に向け, 不同意応答が適する語りの発話の検出可能性とその課題について述べる。

3 検出手法

本研究の目的は, 不同意応答が適する語りの発話の検出の実現可能性を考察することである。本研究では, BERT [13] を用いた単純な手法を採用する。幅広いタスクで有効性が示されている BERT を用いた手法に対して, その検出性能を評価するとともに, その課題を考察する。

本手法では, 語り内の各発話を不同意応答が適

するか否かの2クラスに分類する。分類は、事前学習済みのBERTをfine-tuningしたモデルによって行う。具体的には、分類対象の発話の先頭に[CLS]トークンを、末尾に[SEP]トークンをそれぞれ追加した文字列をモデルに入力する。BERTの最終層の[CLS]トークンの出力を2クラス分類用の出力層で変換して、両クラスに属する確率値を得たのち、確率値が高い方のクラスに分類する。

4 検出実験

4.1 実験概要

不同意応答が適する発話の検出可能性とその課題を考察するために、実験を実施した。述語を中心とした言語的なまとまりである節を語りの発話単位とし、節を不同意応答が適するか否かに分類する手法を実装し、その性能を評価した。

4.2 実験データ

本実験の語りデータには、高齢者のナラティブコーパスJELiCo [25]を用いた。この語りデータには、30名の高齢者による1人約20分の語りの音声収録されている。全高齢者共通の10個の質問に対し、その回答を独話として語るという収録形式が採用されている。節境界解析ツールCBAP [26]によって、語りを11,201個の節に分割し、不同意応答が適するとみなせるものを人手で抽出した。この抽出作業は、著者のうちの1人が行った。抽出作業の結果、全体の2.02%にあたる226個の節が、不同意応答が適する節として抽出された。語りに占める不同意応答が適する節の割合は少なく、不均衡なデータとなっている。

4.3 手法の実装

事前学習済みのBERTをfine-tuningすることで、語り内の節を不同意応答が適するか否かに分類するモデルを実装した。モデルの学習における損失関数は、Cross Entropy Lossとした。実験データ内の節を6:2:2に分割し、それぞれを、学習、開発、テストデータとして、テストデータでの性能を評価した。

事前学習済みのBERTには、huggingfaceのcl-tohoku/bert-base-japanese-v2を用いた。実装には、huggingface/transformersのAutoModelForSequenceClassification, TrainingArguments, Trainerを用いた。バッチサイズを32, エポック数を10, 計算精度に関

するパラメータfp16をTrueとした。その他の設定は上記ライブラリのデフォルトのままとした。

上述のBERTによる手法の他に、入力された節をランダムに分類する以下の2つの手法を実装した。

- **random (even)**: 50%の確率でランダムに分類する手法
- **random (balanced)**: 学習データにおける不同意応答が適する節の割合に従って、ランダムに分類する手法

本実験では、random (balanced) は、1.89%の確率で不同意応答が適すると分類する。

4.4 評価方法

本実験では、適合率、再現率、F値を評価指標とする。不同意応答が適する語りの発話に不同意応答できるだけでなく、それが適さない語りの発話には不同意応答をしないことも重要であると考えられる。そこで、本実験では、不同意応答が適する節と適さない節のそれぞれに対して、適合率、再現率、F値を求める。また、そのマクロ平均も評価に用いる。以降では、不同意応答が適する節を本実験における正例、適さない節を負例とする。正例に対する適合率と再現率は、それぞれ次の式で計算する。

$$\text{適合率} = \frac{\text{正例に正しく分類できた節の数}}{\text{正例に分類した節の数}} \quad (1)$$

$$\text{再現率} = \frac{\text{正例に正しく分類できた節の数}}{\text{正例の節の数}} \quad (2)$$

負例に対しても、同様にして計算する。

4.5 実験結果

表2に、テストデータに対する各手法の評価値を、図2に分類結果に関する混同行列を示す。総合的な評価指標であるマクロ適合率、マクロ再現率、マクロF値のすべてにおいて、BERTを用いた手法は、randomの2つの手法を上回っていた。このことから、幅広いタスクで有効性が示されているBERTに基づく手法は、不同意応答が適する節の検出においても、ある程度有効であるといえる。また、BERTを用いた手法とrandomの各手法との間にマクネマー検定を適用したところ、いずれにおいても有意差が確認できた。

各手法の正例および負例に対する検出性能について述べる。表2の通り、BERTを用いた手法が、正例と負例の両方の評価指標で比較的高い値を記録し

表 2 実験結果

	正例			負例			マクロ平均		
	適合率	再現率	F 値	適合率	再現率	F 値	適合率	再現率	F 値
random (even)	0.024	0.528	0.046	0.977	0.486	0.649	0.501	0.507	0.347
random (balanced)	0.063	0.057	0.059	0.978	0.980	0.979	0.520	0.518	0.519
BERT	0.417	0.189	0.260	0.981	0.994	0.987	0.699	0.591	0.624

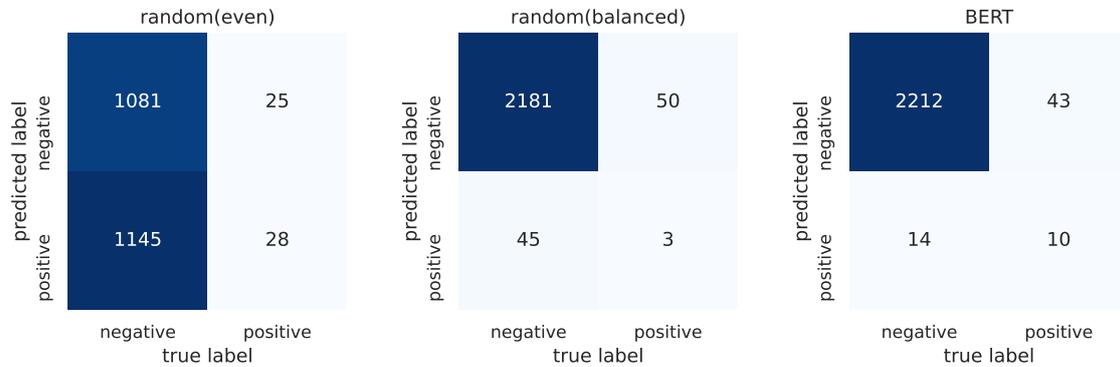


図 2 混同行列

ており、5つの評価指標で random の2つの手法を上回っていた。正例の再現率についてのみ、random (even) が最も高い値となったものの、図 2 の通り、random (even) は約半数もの節を正例であると分類しており、その大部分は誤りであった。

以上の結果から、BERT を用いた手法は、不同意応答が適する節の特徴をある程度捉えられているものといえる。しかし、その性能は十分であるとはいえない。この原因には、本データが正例が少ない不均衡データであったことが挙げられる。正例が少数であるため、正例である不同意応答が適する節の特徴を十分に学習できなかつたものと考えられる。

5 まとめ

本論文では、不同意応答の自動生成に向けた、不同意応答が適する発話の検出について述べた。高齢者の語りデータを用いて、検出実験を実施し、その実現可能性を考察した。実験では、事前学習済みの BERT を用いた手法が、ランダムな手法よりも高い性能を示し、ある程度の性能で検出できることを確認したが、その性能は十分ではなかつた。不同意応答が適する発話は、語りにおける出現が多くない。今後は、こうしたデータの不均衡に起因する問題に対応するために、モデルの学習時におけるサンプリングや class-weight の適用などを検討したい。

謝辞

高齢者のナラティブデータは、奈良先端科学技術大学院大学ソーシャル・コンピューティング研究室から提供いただいた。本研究は、一部、科学研究費補助金 挑戦的研究 (萌芽) (No. 18K19811)、名古屋大学及び JST 科学技術イノベーション創出に向けた大学フェロシップ創設事業 JPMJFS2120 による「名古屋大学融合フロンティアフェロシップ」研究費により実施したものである。

参考文献

- [1] Ministry of Health, Labor and Welfare. Summary report of comprehensive survey of living conditions 2019. https://www.mhlw.go.jp/english/database/db-hss/dl/report_gaikyo_2019.pdf, Jul. 2020.
- [2] Hiroaki Noguchi and Yasuharu Den. Prosody-based detection of the context of backchannel responses. In **Proc. 5th International Conference on Spoken Language Processing**, pp. 487–490, Sydney, Australia, Dec. 1998.
- [3] Nigel Ward and Wataru Tsukahara. Prosodic features which cue back-channel responses in english and Japanese. **J. pragmatics**, Vol. 32, No. 8, pp. 1177–1207, Jul. 2000. DOI:10.1016/S0378-2166(99)00109-5.
- [4] Nicola Cathcart, Jean Carletta, and Ewan Klein. A shallow model for backchannel continuers in spoken dialogue. In **Proc. 10th Conference on European Chapter of the Association for Computational Linguistics**, pp. 51–58, Budapest, Hungary, Apr. 2003. DOI:10.3115/1067807.1067816.
- [5] Shinya Fujie, Kenta Fukushima, and Tetsunori Kobayashi.

- A conversation robot with back-channel feedback function based on linguistic and nonlinguistic information. In **Proc. 2nd International Conference on Autonomous Robots and Agents**, pp. 379–384, Palmerston North, New Zealand, Dec. 2004.
- [6] Norihide Kitaoka, Masashi Takeuchi, Ryota Nishimura, and Seiji Nakagawa. Response timing detection using prosodic and linguistic information for human-friendly spoken dialog systems. *人工知能学会論文誌*, Vol. 20, No. 3, pp. 220–228, Mar. 2005. DOI:10.1527/tjsai.20.220.
- [7] Ronald Poppe, Khiet P Truong, Dennis Reidsma, and Dirk Heylen. backchannel strategies for artificial listeners. In **Proc. 10th International Conference on Intelligent Virtual Agents**, pp. 146–158, Philadelphia, USA, Sep. 2010. DOI:10.1007/978-3-642-15892-6.16.
- [8] Louis-Philippe Morency, Iwan de Kok, and Jonathan Gratch. A probabilistic multimodal approach for predicting listener backchannels. **J. Autonomous Agents and Multi-Agent Systems**, Vol. 20, No. 1, pp. 70–84, Jan. 2010. DOI:10.1007/s10458-009-9092-y.
- [9] Takashi Yamaguchi, Koji Inoue, Koichiro Yoshino, Katsuya Takanashi, Nigel G. Ward, and Tatsuya Kawahara. Analysis and prediction of morphological patterns of backchannels for attentive listening agents. In **Proc. 7th International Workshop on Spoken Dialogue Systems**, pp. 1–12, Saariselkä, Finland, 1 2016.
- [10] Tomohiro Ohno, Yuki Kamiya, and Shigeki Matsubara. Detection of back-channel feedback timings using spoken dialogue corpus. **IEICE Trans. Fundamentals (Japanese Edition)**, Vol. J100-A, No. 1, pp. 53–65, 1 2017.
- [11] Robin Ruede, Markus Müller, Sebastian Stüker, and Alex Waibel. Enhancing backchannel prediction using word embeddings. In **Proc. 18th Annual Conference of the International Speech Communication Association**, pp. 879–883, Stockholm, Sweden, Aug. 2017. DOI:10.21437/Interspeech.2017-1606.
- [12] Jin Yea Jang, San Kim, Minyoung Jung, Saim Shin, and Gahgene Gweon. BPM_MT: Enhanced backchannel prediction model using multi-task learning. In **Proc. 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 3447–3452, Punta Cana, Dominican Republic (Online), Nov. 2021. DOI:10.18653/v1/2021.emnlp-main.277.
- [13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-2019)**, pp. 4171–4186, Minneapolis, USA, 2019.
- [14] 日本語記述文法研究会. 現代日本語文法 7. くろしお出版, 2009.
- [15] Yuka Kobayashi, Daisuke Yamamoto, Toshiyuki Koga, Sachie Yokoyama, and Miwako Doi. Design targeting voice interface robot capable of active listening. In **Proc. 5th ACM/IEEE International Conference on Human-Robot Interaction**, pp. 161–162, Osaka, Japan, Mar. 2010.
- [16] Toyomi Meguro, Yasuhiro Minami, Ryuichiro Higashinaka, and Kohji Dohsaka. Evaluation of listening-oriented dialogue control rules based on the analysis of HMMs. In **Proc. 12th Annual Conference of the International Speech Communication Association**, pp. 809–812, Florence, Italy, Aug. 2011.
- [17] Kazuya Shitaoka, Ryoko Tokuhisa, Takayoshi Yoshimura, Hiroyuki Hoshino, and Narimasa Watanabe. Active listening system for a conversation robot. **Journal of Natural Language Processing (Japanese Edition)**, Vol. 24, No. 1, pp. 3–47, 2017. DOI: 10.5715/jnlp.24.3.
- [18] Koji Inoue, Divesh Lala, Kenta Yamamoto, Shizuka Nakamura, Katsuya Takanashi, and Tatsuya Kawahara. An attentive listening system with android ERICA: comparison of autonomous and WOZ interactions. In **Proc. 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue**, pp. 118–127, 1st virtual meeting, Jul. 2020.
- [19] 村田匡輝, 大野誠寛, 松原茂樹. 系列変換モデルに基づく傾聴的な応答表現の生成. 言語処理学会第 24 回年次大会発表論文集, pp. 821–824, 岡山コンベンションセンター, Mar. 2018.
- [20] 伊藤滉一朗, 村田匡輝, 大野誠寛, 松原茂樹. 傾聴を示す応答で繰り返される語りの語句の検出. 言語処理学会第 25 回年次大会発表論文集, pp. 1316–1319, 名古屋大学東山キャンパス, Mar. 2019.
- [21] 川本稔己, 長谷川駿, 上垣外英剛, 船越孝太郎, 奥村学. 傾聴の応答で繰り返される語句の検出性能の向上. 言語処理学会第 27 回年次大会発表論文集, pp. 1580–1584, 北九州国際会議場, Mar. 2021.
- [22] 田原俊一, 松本一則, 服部元. オウム返し応答を生成する対話システム. FIT2021 (第 20 回情報科学技術フォーラム), 第 2 巻, pp. 465–466, オンライン, 2021.
- [23] Toshiaki Kawamoto, Hidetaka Kamigaito, Kotaro Funakoshi, and Manabu Okumura. Generating repetitions with appropriate repeated words. In **Proc. 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 852–859, Seattle, United States, Jul. 2022. DOI:10.18653/v1/2022.naacl-main.62.
- [24] Koichiro Ito, Masaki Murata, Tomohiro Ohno, and Shigeki Matsubara. Construction of responsive utterance corpus for attentive listening response production. In **Proc. 13th Language Resources and Evaluation Conference**, pp. 7244–7252, Marseille, France, June 2022.
- [25] Eiji Aramaki. Japanese elder’s language index corpus v2. https://figshare.com/articles/dataset/Japanese_Elder_s_Language_Index_Corpus_v2/2082706/1, Nov. 2016.
- [26] Hideki Kashioka and Takehiko Maruyama. Segmentation of semantic units in Japanese monologues. In **Proceedings of the International Conference on Speech and Language Technology & Oriental-COCOSDA 2004 (ICSLT-O-COCOSDA 2004)**, pp. 87–92, Delhi, India, Nov. 2004.